The knowledge network

Cabo, P.G.

Publication date:
1997

CHAPTER 6
A KEYWORD ANALYSIS OF JOINT BIOTECHNOLOGY PROJECTS

INTRODUCTION

Biotechnology is one of the areas in which technological cooperation between firms is most prevalent. It is also an area in which universities and research institutes play an important role. In fact, science and technology are highly interwoven in biotechnology (Kenney 1986, Orsenigo 1986)[11]. The main focus of attention in this chapter is on the methodology of technology studies. To assess this methodology I will apply it to several research questions: how can we characterize the content of the technology in R&D projects? How can the activities of universities and firms be described and compared? These questions are especially relevant for high-tech fields because of the ‘strategic’ nature of R&D in these fields. This means that results from basic science have immediate commercial effects, and that science based firms are a significant part of the research community.

The empirical basis for this paper are projects from the ECLAIR programme (the European Collaborative Linkage of Agriculture and Industry through Research). The size of this program, 42 projects, permits an illustration of the detailed analysis that I advocate, but, at the same time, there are a relative large number of joint projects of firms and research organisations. Furthermore, the public databases from the European Commission provide excellent documentation about R&D activities.

The methodology that I develop is an extension of the work reported in Chapter

[11] See also Chapter 8.

5 where content analysis of project descriptions is used in order to describe the content of R&D projects. In this chapter I assume that the type of innovation in the projects is determined by analyzing the keywords that each participant uses to characterize their activities in the project. One advantage of using keywords is that patterns of keywords reflect the content of the research, or ‘problem field’, as perceived by researchers. To make typologies of innovations is always difficult: typologies are either to general or the social scientist has not enough expert knowledge about the type of R&D. The use of keywords removes these problems. Another advantage of keywords is that they are standardized, as opposed to title words or abstract words. If researchers have to choose from a fixed pre-determined list, it becomes easy to compare their patterns of keywords.

First, I combine all key-word files for firms, research institutes, and universities. My goal is to analyze the differences between these three organizations with respect to three classes of keywords: result types, subject descriptors, and market applications. Since the projects are international, I also check for differences between countries. Secondly, we take a more detailed look at five joint projects in order to examine, by using correspondence analysis, the ‘problem structure’ of the cooperation between firms, research institutes and universities.

ECLAIR

The European Collaborative Linkage of Agriculture and Industry through Research (ECLAIR) programme was launched under the second Framework programme, subactivity 4.2. : "Agro-industrial technologies" for agricultural and industrial advancement through technological development, using biotechnology and other advanced technologies. This programme was aimed at improving links between agriculture and industry by taking advantage of Europe’s strength in biotechnology and the life sciences. The programme can be characterized as joint R&D or as cooperative research. Under this programme 42 projects were initiated, starting in 1988 and finishing in 1993. In 1994, 100 files described the results obtained by participants. Under this programme, research institutes and/or universities should participate in a group together with one or more firms. Most of the time the research groups were initiated by one participant who contacted other participants with appropriate scientific skills. The research carried out in the ECLAIR programme is pre-competitive in nature, in that any products or processes developed in the programme will require a further two to three years development before commercialization. One of the main priorities of the programme is to encourage the transfer of technology in the areas where it will have the most practical and economic benefit. In the programme, three areas of research are distinguished:

- Evaluation trials and production of new species or organisms: test trials at the appropriate scale and under various conditions into novel or modified species or organisms (plants, livestock and other);

- Industrial products and services: more precise and effective inputs to agriculture, more precise and effective extraction, transformation and production processes;

- Integrated approaches: development of systems for harvesting the whole of a crop, studies and development projects for the integrated use of new technologies.

Data about ECLAIR was obtained from ECHO (the European Commission Host Organization in Luxembourg). From the PROJECTS database I obtained information about the participants in ECLAIR projects. From the RESULTS database I downloaded information about results of the projects. The BIOREP database was used for additional information.

Our main variables, organizations and results, are operationalized in the following way:

Organizations are categorized as firms, universities, research institutes, or other.

For the *results* of R&D projects I use three items that originate directly from the database.

- *Type of result*: this item has two discriminating categories: process or prototype, and methodology, skill, know how.
- *Subject discriptors*: this is a set of keywords that indicate the disciplines covered by the result.
- *Market application*: this is a set of keywords that indicate the commercial and industrial sectors in which the result may have commercial potential.

Together, type of results, subjects, and markets define what I call type of innovation.

Each project has a file that contains the names of the participating organizations. Each project has also a number of result records. Each result record contains the name of the organization that is most responsible for the result. The project and result files can be linked by a common code.

In the 42 project descriptions firms were identified 114 times, research institutes 136 times, and universities 76 times[12]. This means that a typical project consists of three firms, three research institutes and two universities. Actually, most projects (24) contain at least one organization from each category. Seven projects contain only firms and research institutes and seven projects contain only firms and universities. There are two projects with only research institutes and

[12] 11 participants are classified as "other", 14 participants are undefinable. This makes a total of 351 participants.

universities and two projects with only firms[13]. In 57% of the projects a firm is the main participant (research institute 24%, university 19%).

Most projects have several result files and there can be different types of results from one project. Surprisingly, while research institutes are the most common organization, they have the smallest number of results (26%). Most results are assigned to industry (45%).

RESULT TYPE, SUBJECT DESCRIPTORS, AND MARKET APPLICATIONS

Table 1 shows that the three types of organizations have about the same types of results. A relative comparison between industry and research institutes shows that institutes have significantly more methods results. The other comparisons are not significantly different (see note 11).

If research institutes and universities were to be combined in one ’research’ category, the chi-square would be 4.16 with p=0.041.

Table 1 Result type per organization (number of records)

Prime contractor	Process, prototype	Methodology, skill, know-how	Total
Industry	12	33	45
Research Institute	2	24	29
University	4	25	26
Total	18	82	100

Chi-sq= 4.509, p= 0.105

[13] These last two combinations can appear, because of the total of undefinable participants. Except for one project, the category "other" is always joined with the cooperation between firm, university and research institute together.

When industry is recorded as prime contractor for the result, it seems that the sector Biochemistry and Microbiology are focused on, for research institutes the sector Bioengineering seems important, university's main concern lies within Genetics. Regarding the different patterns of subject descriptors, there's not a significant correlation between the three participants[14].

Table 2 Subject descriptors per type of organization (percentages)

Subject descriptor	**Industry** 45=100%	**Research institute** 26=100%	**University** 29=100%
Agriculture	**.53**	**.54**	**.45**
Biochemistry	**.28**	.25	.31
Bioengineering	.24	**.46**	.21
Biotechnology	**.44**	.25	**.41**
Botany	.13	.21	.28
Cell biology	.18	.25	.28
Food science	.16	**.29**	.17
Genetics	.11	**.29**	**.41**
Microbiology	**.28**	.13	.24

In addition to the disciplinary background, an overview can be given for the main commercial and industrial sectors in which the result may have commercial potential.

[14]Correlations:	Industry	Research Institute	University
Industry	1.00	.4613	.5566
Research Institute		1.00	.2946
University			1.00

Table 3 Market applications per type of organization (percentages)

Market application	**Industry** 45=100%	**Research institute** 26=100%	**University** 29=100%
Animal, plant breeding	**.27**	**.46**	.31
Biotechnology	**.60**	**.67**	**.69**
Farming	**.36**	.17	.14
Fermentation	**.27**	.04	.17
Genetic, protein engineering	.16	.29	**.34**
Horticulture	.11	**.33**	**.48**

For Industry, the Farming sector is prominent; Animal, plant breeding is an important sector for Industry as well as for Research Institutes. For Universities, the main sector is Horticulture.

NATIONAL DIFFERENCES

In this section I want to test whether differences between the results of projects exist not only at the level of organizations but also at the level of nations. First, I will present some general data about this international dimension.

Eleven countries participate in ECLAIR. Table 7 shows how the total number of organizations, projects and results are distributed among countries. It can be seen that France has the highest share in all categories.

Table 4 Involvement in ECLAIR by country (percentages)

Country	Number of organizations 334=100%	Number of projects 42=100%	Number of results 100=100%
Belgium	.09	.36	.14
Denmark	.04	.19	.04
France	.28	.67	.27
Germany	.06	.24	.07
Greece	.04	.19	.01
Ireland	.05	.26	.09
Italy	.12	.36	.05
Netherlands	.06	.26	.07
Portugal	.04	.26	.01
Spain	.11	.50	.05
United Kingdom	.11	.55	.20

The number of countries per project ranges from two to seven, with an average of four. Both absolutely and relatively, most (40%) joint projects are between France and Spain. Cooperation between France and Belgium, France and Italy, Italy and Spain, and Denmark and the Netherlands is also relatively frequent.

Since France and the United Kingdom have a large number of result records I will make a comparison between these two countries.

Table 5 Type of results for selected countries

Country	Process, Prototype	Methodology, skill, know-how	Total
France	5	22	27
UK	5	15	20
Total	10	37	47

Chi-sq= 0.288, p=0.591

Although the UK have relatively more process/prototype results than France there is no significant difference between the two countries.

As we did with the organizations, we determined the most frequent subject descriptors (table 9) and market applications (table 10) for France and the UK. As a rule, a key-word is included if it appears at least twice in one country. "Bio-technology" and "agriculture" are considered redundant words and are not included.

Table 6 Subject descriptors for selected countries (percentages)

Subject descriptor	France (27=100%)	United Kingdom (20=100%)
Aquaculture	.07	.15
Biochemistry	.22	***.30***
Bioengineering	***.30***	***.35***
Botany	.11	.20
Brewing	.07	0
Cell biology	.15	.15
Chemical engineering	.22	.05
Chemicals	.15	.05
Food science	***.33***	.20
Genetics	.22	.25
Horticulture	.11	.25
Immunology	.11	.05
Microbiology	***.30***	.25
Nutrition	.07	0
Pathology	.07	0
Pharmacy, pharmacology	.07	.05
Physiology	.15	.10
Pollution	0	.10
Polymers	0	.20

"Chemical engineering", "food science", and "chemicals" are the typical French words. "Polymers", "horticulture", and "pollution" are typical British words. The correlation between the patterns is 0.56 and not significant. This indicates that France and the UK have results on different subjects.

Table 7 Market applications of selected countries

Market application	France (27=100%)	UK (20=100%)
Agricultural chemicals	0	***.20***
Animal, plant breeding	***.26***	.20
Biological pest control	.11	.15
Farming	.22	***.30***
Fermentation	***.30***	.10
Fishing industry, fish farming	.07	.05
Food products	***.33***	.10
Foods, drinks	.15	.10
Genetics, protein engineering	.11	***.20***
Horticulture	.18	***.20***
Industrial chemicals	.22	.20
Measurement systems	.07	.05
Nutritional products	.18	.05
Organic agriculture	.07	.20
Paper, paper products	.07	.05
Pharmaceutical products	.11	.05
Physiological monitoring	.11	.10

Again, the two countries have highly dissimilar patterns of market applications (correlation is 0.078 and not significant). France specializes in "food products (animal, human)", "fermentation", and "nutritional products". British markets are: "agricultural chemicals", "genetic, protein engineering", and "organic agriculture".

DISCUSSION

Although the data do not allow us to compare the innovations of all countries, the comparison of France and the United Kingdom indicate that specialization is not only a matter of differences between organizations but also of differences between nations.

When I examined the type of innovations I found that research institutes have the smallest number of result records. So, although these organizations participate a lot, their role could be characterized as ’supportive’. An answer to the question whether the nature of industrial R&D is similar to public sector R&D depends on the specific dimension of innovation. I defined innovation as a combination of subject, market, and type. The last dimension, which differentiates between process/prototype and knowledge/skill/know-how, is not very strongly affected by the organizational dimensions. First of all, almost all results are in the methodology category. This is in line with the objectives of ECLAIR, and more general EC policy, which favour precompetitive R&D. As could be expected, firms have relatively more prototype results. This is significant if we compare firms to research institutes, but not significant if they are compared to universities. Interestingly, when we looked at projects consisting of all three types of organizations, the number of prototype results is even smaller. This may indicate an effect of the sheer number of participants: the more participants the less likely are applied results. A reason for this could be that a large number of organizations, especially firms, causes appropriability problems.

The subject and market indicators show that there are large differences between industry, research institutes and universities, and various combinations of these actors. Universities are clearly focused on the more strategic fundamental areas: genetics, and protein engineering. These results support the idea that there is some complementarity of the different types of organizations. This would be a rationale for cooperation. Of course, this does not exclude complementarity within a category: two firms (for instance a large pharmaceutical firm and a small new biotechnology firm) can have complementary assets as well. An answer to this question could be obtained by comparing the innovations per project. This will be discussed in the next section.

CORRESPONDENCE ANALYSIS OF JOINT PROJECTS

In order to analyze differences between organizations in one project I selected the projects that contain result files from at least one university, one research institute, and one firm. 5 cases satisfied this criterion. In each case I made a matrix in which the organizations are the columns and the keywords, from the subject and market categories, are the rows. Since the result type is almost always 'methodology, skill, know-how', this category has been left out. Correspondence analysis is a method for the graphical representation of the rows and columns in one map (Greenacre 1993, Welter and Romney 1990). This enables us to determine, at the same time, the differences between the organizations, and the key-words that account for these differences. Applied to joint R&D projects, this technique gives an impression of the task and organizational complexity of alliances (Killing 1988)[15]. I used the correspondence analysis module of UCINET IV software (Borgatti et al. 1992) with the default options. I also used UCINET to compute the Euclidean distances between organizations, as a relative dissimilarity measure.

Case 1: Novel antifungal proteins: applications in crop protection

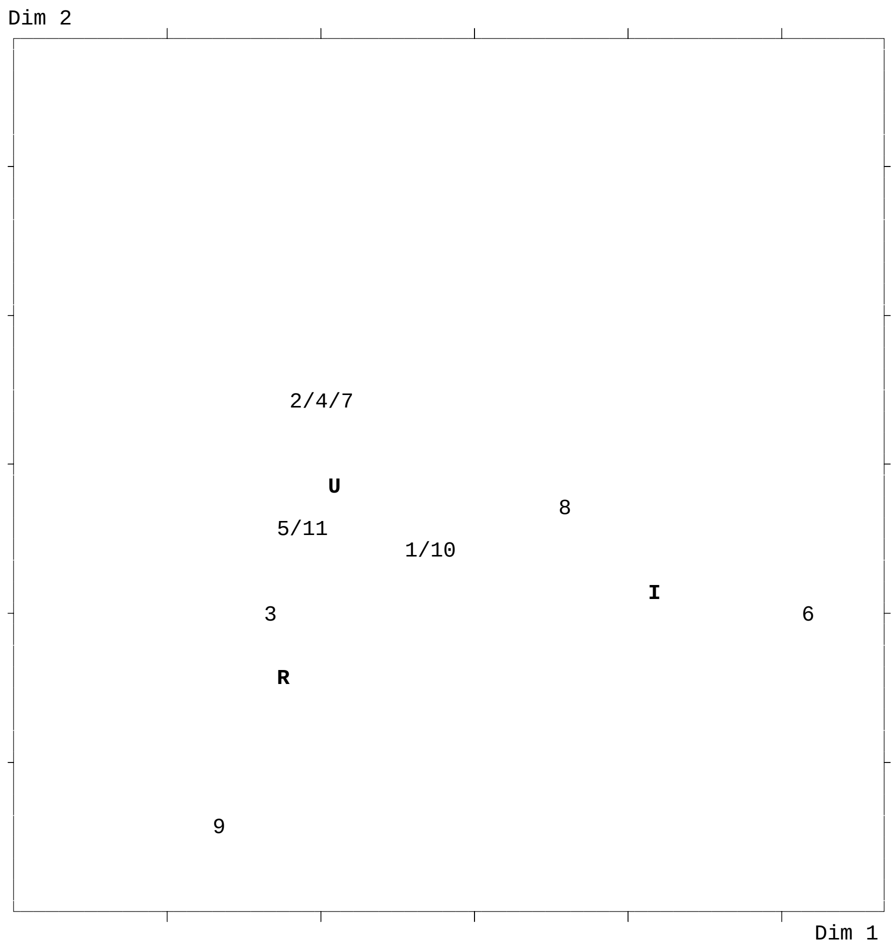
In this project there are two result files from a university (U, the main participant) one from a firm (I) and one from a research institute (R)

```
SINGULAR VALUES

  FACTOR  VALUE PERCENT   CUM %   RATIO
 ------- ------ ------- ------- -------
      1:  0.634    58.1    58.1   1.385
      2:  0.458    41.9   100.0
 ======= ====== ======= ======= =======
          1.091   100.0
```

[15] See also Chapter 3.

Scatterplot of Scores on First Two Non-Trivial Factors



Keywords:

Subjects:
1 Agriculture
2 Biochemistry
3 Botany
4 Genetics

Markets:
5 Animal, plant breeding
6 Biological pest control
7 Farming
8 Genetic, protein engineering
9 Measurement systems

Subjects/markets:
10 Biotechnology
11 Horticulture

DISSIMILARITIES

	U	R	I
U	0.00	1.10	1.65
R	1.10	0.00	1.90
I	1.65	1.90	0.00

Case 2: The development of environmentally safe pest control systems for the European olive

In this projects there are result files from four research institutes (R1 to R4), one firm (I), and one university (U) (the main participant).

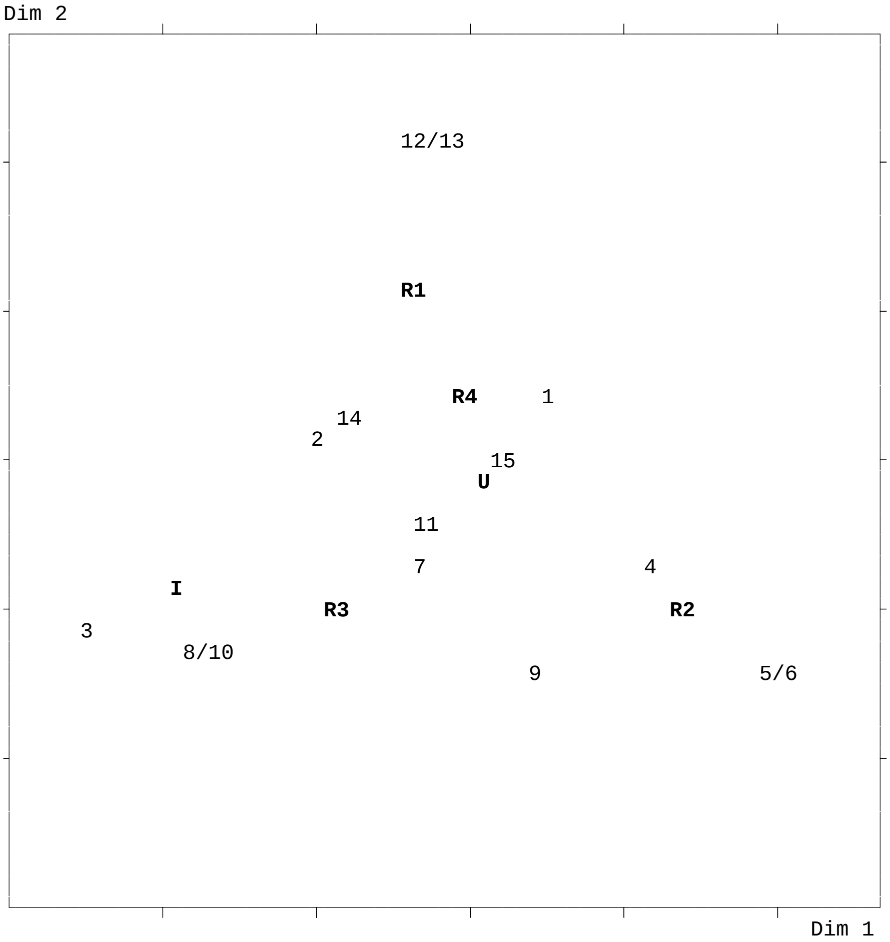
SINGULAR VALUES

FACTOR	VALUE	PERCENT	CUM %	RATIO
1:	0.704	27.6	27.6	1.128
2:	0.624	24.4	52.0	1.389
3:	0.449	17.6	69.6	1.085
4:	0.414	16.2	85.8	1.138
5:	0.364	14.2	100.0	
	2.554	100.0		

DISSIMILARITIES

	R1	R2	R3	R4	U	I
R1	0.00	2.00	1.65	0.63	1.09	1.84
R2	2.00	0.00	1.50	1.39	0.99	2.22
R3	1.65	1.50	0.00	1.17	0.93	0.73
R4	0.63	1.39	1.17	0.00	0.45	1.57
U	1.09	0.99	0.93	0.45	0.00	1.51
I	1.84	2.22	0.73	1.57	1.51	0.00

Scatterplot of Scores on First Two Non-Trivial Factors



Subjects:
1 Agriculture
2 Biochemistry
3 Bioengineering
4 Biological interactions
5 Expert systems
6 Information analysis
7 Microbiology
8 Toxicology
9 Zoology

Markets:
10 Agricultural chemicals
11 Biological pest control
12 Organic agriculture
13 Physiological monitoring

Subjects/Markets:
14 Biotechnology
15 Horticulture

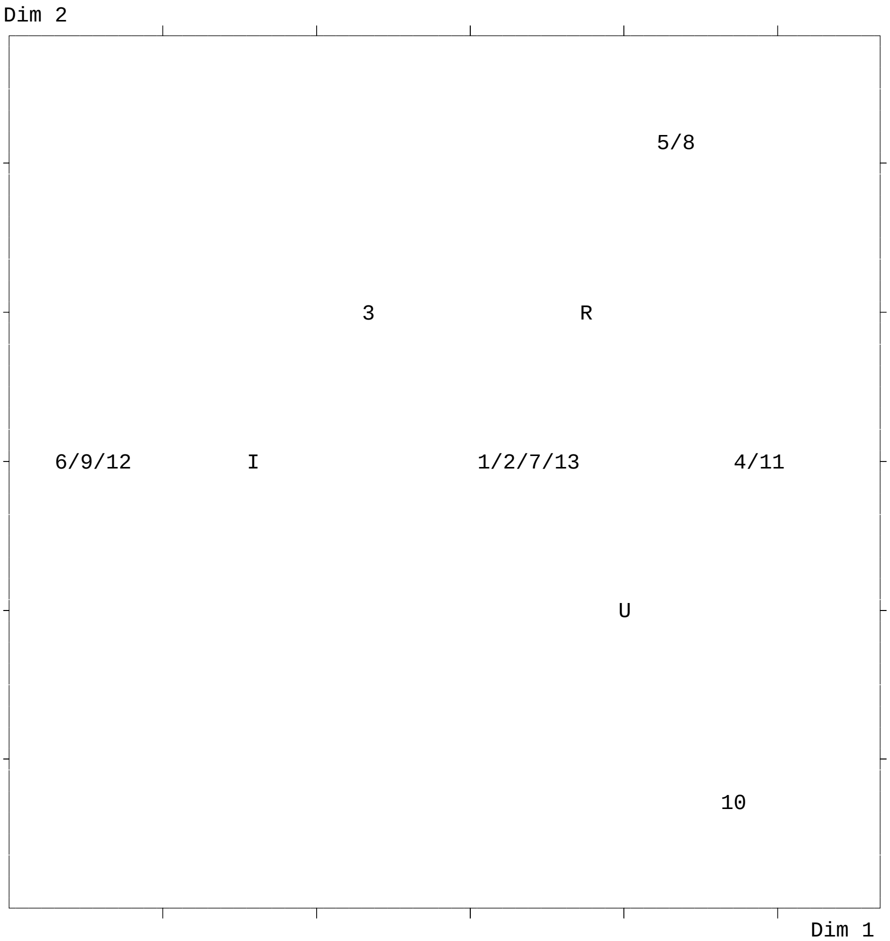
Case 3: The development, refinement and commercialization of a biotechnology based on the in vitro production, sex determination, freeze-storage and transfer of bovine embryos.

In this project there are result files from a firm (I, the main participant), a research institute (R), and a university (U).

SINGULAR VALUES

FACTOR	VALUE	PERCENT	CUM %	RATIO
1:	0.638	58.2	58.2	1.390
2:	0.459	41.8	100.0	
	1.097	100.0		

Scatterplot of Scores on First Two Non-Trivial Factors



Subject:
1 Agriculture
2 Animal husbandry
3 Bioengineering
4 Cell biology
5 Physiology
6 Veterinary science

Market:
7 Animal, plant breeding
8 Farming
9 Genetic, protein engineering
10 Measurement systems
11 Physiological monitoring
12 Veterinary equipment

Subject/Market:
13 Biotechnology

DISSIMILARITIES

	U	R	I
U	0.00	1.13	1.52
R	1.13	0.00	1.42
I	1.52	1.42	0.00

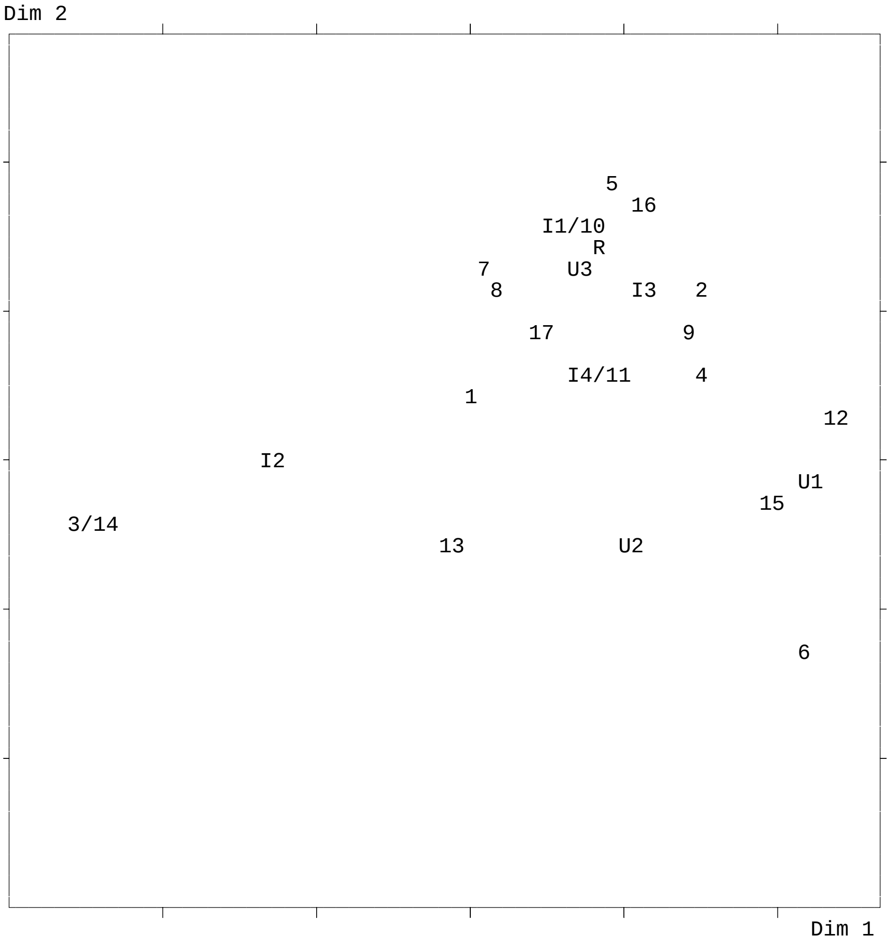
Case 4: Biological inoculants for seed/plant establishment

This project has result files from four firms (I1 to I4, I4 is the main participant), three universities (U1 to U3), and one research institute (R).

SINGULAR VALUES

FACTOR	VALUE	PERCENT	CUM %	RATIO
1:	0.612	21.8	21.8	1.177
2:	0.520	18.5	40.2	1.108
3:	0.469	16.7	56.9	1.157
4:	0.405	14.4	71.3	1.224
5:	0.331	11.8	83.1	1.261
6:	0.262	9.3	92.4	1.228
7:	0.214	7.6	100.0	
	2.812	100.0		

Scatterplot of Scores on First Two Non-Trivial Factors



Subject:
1 Agriculture
2 Biochemistry
3 Bioengineering
4 Botany
5 Cell biology
6 Genetics
7 Microbiology
8 Pathology
9 Physiology

Markets:
10 Animal, plant breeding
11 Biological pest control
12 Biotechnology
13 Farming
14 Fermentation
15 Genetic, protein engineering
16 Physiological monitoring

Subjects/markets
17 Horticulture

DISSIMILARITIES

	I1	I2	I3	I4	R	U1	U2	U3
I1	0.00	1.59	0.53	0.68	0.29	1.65	1.54	0.23
I2	1.59	0.00	1.79	1.38	1.73	2.33	1.60	1.56
I3	0.53	1.79	0.00	0.48	0.25	1.14	1.20	0.33
I4	0.68	1.38	0.48	0.00	0.56	1.13	0.86	0.47
R	0.29	1.73	0.25	0.56	0.00	1.38	1.38	0.17
U1	1.65	2.33	1.14	1.13	1.38	0.00	0.85	1.42
U2	1.54	1.60	1.20	0.86	1.38	0.85	0.00	1.32
U3	0.23	1.56	0.33	0.47	0.17	1.42	1.32	0.00

Case 5: Improvement of yield and feed conversion in salmonids and maricultured fish

This project has result files from one firm (I, the main participant), one research institute (R), and two universities (U1 and U2).

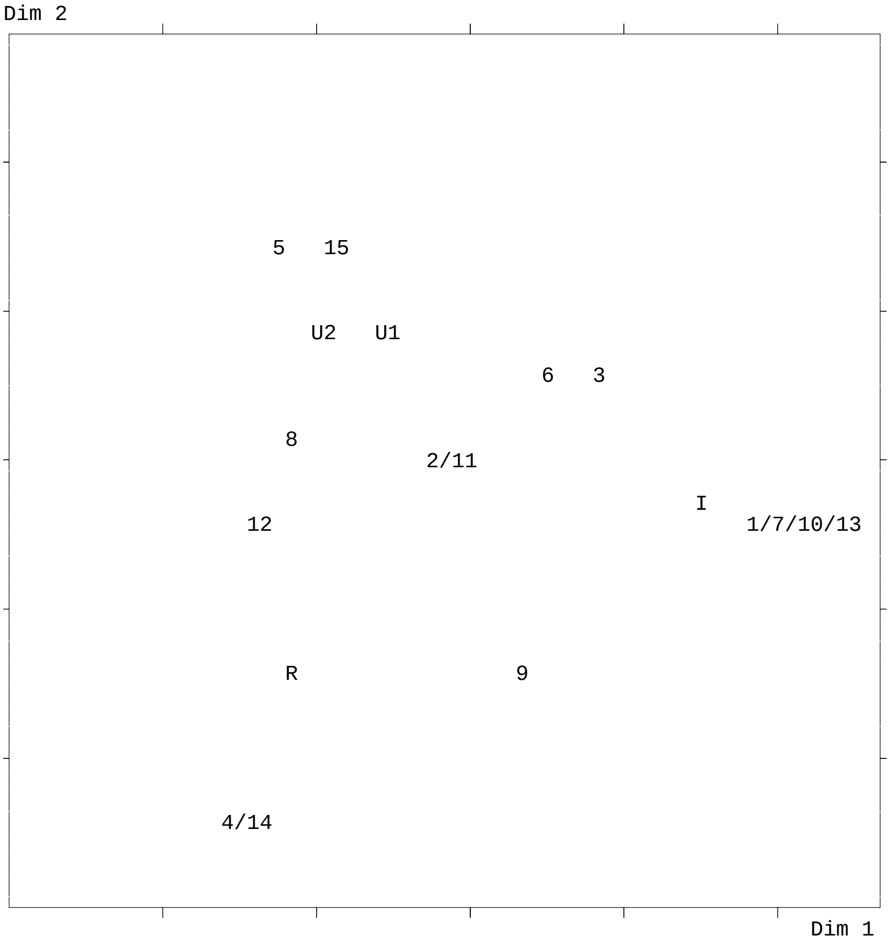
SINGULAR VALUES

FACTOR	VALUE	PERCENT	CUM %	RATIO
1:	0.718	41.2	41.2	1.205
2:	0.596	34.2	75.3	1.386
3:	0.430	24.7	100.0	
	1.745	100.0		

DISSIMILARITIES

	U1	U2	R	I
U1	0.00	0.27	1.52	1.45
U2	0.27	0.00	1.51	1.70
R	1.52	1.51	0.00	1.83
I	1.45	1.70	1.83	0.00

Scatterplot of Scores on First Two Non-Trivial Factors



Subjects:
1 Agriculture
2 Aquaculture
3 Biochemistry
4 Bioengineering
5 Biotechnology
6 Food science
7 Marine science
8 Physiology

Markets:
9 Animal, plant breeding
10 Farming
11 Fishing industry, fish farming
12 Food products (animal, human)
13 Foods, drinks
14 Measurement systems
15 Physiological monitoring

DISCUSSION

The first case, like case 3, is a simple one because there are only three actors involved. This means that the overlap between organizations in terms of keywords could also be represented in a Venn diagram. In such a diagram, overlapping circles, representing organizations, would show the keywords that are common to all participants, keywords that are common to two participants, and keywords that are specific for one organization. If there are more than three actors (such as in cases 2, 4, and 5) the picture becomes to complicated and correspondence analysis will help. Also, if keywords occur more than once with one actor (this happens with the university in case 1) correspondence analysis will take account of this 'weight'.

The singular values that I report with each plot indicate the amount of variation that each axis accounts for. Thus, in case 1, the horizontal axis accounts for 58%, and the vertical axis for 42% of the variation in the data. If we take a look at the plot we see that 'agriculture' and 'biotechnology' are in the center of the space delimited by the three organizations. This means that these words occur with each of the three organizations. In this case, this is not surprising since these words are general for the program. We can also see wich words are organization specific: 'biochemistry', genetics', and 'farming' are specific for the university, 'biologiacal pest control' is specific for the firm, and 'measurement systems' is specific for the research institute. The other words are 'inbetween' organizations.

We suggest two measures for the degree of commonality or specificity of a *problem field* (the set of keywords). A problem field is *centrifugal* when there is a large number of specific words (in case 1, 36%), and a problem field is *centripetal* when there is a large number of common words (18% in case 1).

The dissimilarity scores, at the end of each case, measure the exact difference between organizations in the two-dimensional plot. So, the largest difference exists between firm and research institute (1.90), and the smallest difference is between university and research institute (1.10).

How can the cases be characterized in general terms? One of the research questions I mentioned in the introduction was about the specialization of different types of organizations: is public sector research different from private sector research? Cases 1, 3, and 5 would support this view. Universities and research centres are on one side of the map, firms are on the other side. However, this does not mean that there is absence of overlap between the keyword domains of the organizations (problem fields are never completely centrifugal). Also, there are important differences between research institutes and universities. Cases 2 and 4 do not support the differentiation hypothesis. Case 2 shows a clique of research institutes with a university in the centre, and a firm attached to this clique. Nevertheless, one research institute (R3) is closer to the firm than to all other organizations. In case 4, universities, firms and the research institute all mix[16].

One hypothesis that came up when I plotted the data is about the role of the main participant. Does the fact that an organization is project leader has a greater effect on its position in the 'problem domain' (the set of keywords) than its organizational origin? In cases 1, 2, and 4 we find the main participant in the centre of the plot. So, being project leader means that one occupies an 'average' position in the problem field, linking all other participants together. However, cases 3 and 5 do not support this view. The firms that are project leaders in these cases no not have the most central position (these are occupied by R and U1 respectively). Maybe, the hypothesis should be that public sector organizations tend to occupy central positions in problem fields.

Two other sources of variation were considered and found not to be significant:

---

[16] In case 4, the two-dimensional plot is problematic because it accounts for only 40% of the variation. A three dimensional plot would account for 57%. This means that the two-dimensional distances have to be interpreted with care.

the difference between market words and subject words, and the effect of nationality of the participants.

CONCLUSION

This paper showed how keyword analysis can be used for the study of technological innovation in biotechnology at two levels of aggregation: projects and programs (as collections of projects). Keyword analysis, whether as word patterns (Braam 1991) or co-word structures (Rip and Courtial 1984) is not new but its application to records in which participants of R&D projects describe their activities has been developed more recently (see chapter 6). To my knowledge it is the first time such an analysis has been performed at the level of single R&D projects. The analysis of ECLAIR showed how the innovation activity of public and private sector organizations can be compared, how degrees of similarity an be computed, and how differences can be a qualified.